https://doi.org/10.2352/ISSN.2470-1173.2021.9.IQSP-263

Quality is in the Salient Region of the Image

Meisam Jamshidi Seikavandi*†, *Seyed Ali Amirshahi*
† ***Khaje Nasir University of Technology, Tehran, Iran***
* ***The Norwegian Colour and Visual Computing Laboratory, Norwegian University of Science and Technology, Gjøvik, Norway***
meisamjam@gmail.com, s.ali.amirshahi@ntnu.no

Abstract

The ultimate goal in any proposed Image Quality Metrics (IQMs) is to accurately predict the subjective quality scores given by observers. In the case of most IQMs the quality score is calculated by pooling the quality scores from what is referred to as a quality map of an image. While different pooling methods have been proposed, most such approaches use various types of a weighting average over the quality map to calculate the image quality score. One such approach is to use saliency maps as a weighting factor in our pooling process. Such an approach will result in giving a higher weight to the salient regions of the image. In this work we study if we can evaluate the quality of an image by only calculating the quality of the most salient region in the image. Such an approach could possibly reduce the computational time and power needed for image quality assessment. Results show that in most cases, depending on the saliency calculation method used, we can improve the accuracy of IQMs by simply calculating the quality of a region in the image which covers as low as 20% of the salient energy.

Introduction

Anyone who works on aesthetic or general quality evaluation of images has come across some version of the phrase "beauty is in the eye of the beholder" [1]. In fact, many studies have either based their work on, or came up with a conclusion which could be linked to this term. While this term can be interpreted in different ways, when it comes to images, we can all agree that it is the observer who decides about the quality of an image.

For decades researchers have tried to introduce an Image Quality Metric (IQM) which is able to objectively evaluate the quality of images [2]. Such a focus has resulted in many different IQMs, where some are able to evaluate the quality of images that closely resemble that of a human observer. In fact, the main goal in objective image quality assessment is to introduce an IQM which has the highest correlation possible with the perceived subjective image quality evaluation done by human observers. Various IQMs have shown a significantly high correlation to subjective scores given by observers, nevertheless, different parameters such as run-time performance and memory requirements are seen as challenges that need to be addressed when proposing new IQMs [3].

Over the years researchers have used different approaches to evaluate the quality of images [4]. This includes but is not limited to structural similarity [5], spatial extensions of color difference formulae [6], human perception [7], machine learning [8, 9], etc. [2, 10]. While saliency approaches have been widely used in the quality assessment of images and videos [11, 12, 13] it has mostly been used for giving a higher weight to more salient regions in images and videos. In their work Alers et al. [14] have taken a different approach in where they calculate two different quality scores one for the salient region in the image and another for the its background. They have then studied the relationship between the two introduced objective quality scores and subjective scores given by observers. Using saliency calculations, in this work, we take a step further and investigate if we can evaluate the quality of an image by just evaluating the quality of the most salient region in the image. We keep the ratio and pixels' combination in an image untouched to provide a method that is applicable to all IQMs. Simply said, we investigate if a small but highly salient region in the image can be used as a good representative of the overall quality of the image. With the increase in the size and resolution of images and videos this work can be seen as a first step to investigate how the computational time and power for calculating the quality of images and videos could be reduced while at the same time not deteriorating the performance of the proposed image and video quality metrics in a dramatic way. This study could also be the first steps for introducing the concept of personalized IQM which has previously been introduced in the case of evaluating the aesthetic quality of paintings and images [15].

In the rest of the paper we first introduce the proposed approach, in the next section experimental results are discussed, and finally a conclusion of the findings are presented.

Proposed Approach

For a long time eye tracking devices and saliency detection techniques have been widely used in different computer vision and image processing applications [17]. This is especially the case in aesthetic and general quality assessment in which saliency maps have been used to evaluate the overall quality in photographs [18] and paintings [1] or for detecting the existence of the rule of thirds in the same mentioned media [19, 20, 21]. In the case of general quality, saliency maps have been used in image [22] and video [11] quality assessment. As mentioned earlier, in the case of such applications, the saliency maps are used as a weighting factor in which a higher weight is given to more salient regions in the images and videos. This approach assumes that the salient region in an image correspond to regions that observers pay more attention to and so when evaluating the quality of an image their quality has a higher importance to the observer. While initially such reasoning is justified, using different eye tracking experiments, it has been shown that given the task to evaluate the quality of images, observers only take one or two seconds to decide on the quality of the image [23]. Naturally, such a finding could put doubts on the accuracy of using saliency maps as a weighting factor for image quality assessment. In other words, we can confidently assume that the few seconds used to evaluate the quality of an image would not be enough to accurately observe the entire image in detail. Keeping this issue in mind, in this work we study if we can simply evaluate the quality of an image using only the most salient region of the reference image without taking the rest of the image into account. For this, the following steps are taken:

(a) (b) GBVS saliency map (c) 20% saliency (d) 40% saliency (e) 50% saliency (f) 60% saliency (g) 80% saliency

(h) (i) GBVS saliency map (j) 20% saliency (k) 40% saliency (l) 50% saliency (m) 60% saliency (n) 80% saliency

Figure 1. *Sample images from the CSIQ database [4] with the corresponding GBVS saliency [16] maps and the cropped images. While in each case we only calculate the image quality for the cropped regions for a better understanding of the position of the salient region in the image borders have been added to the figures to show the location of the salient region with regards to the image.*

1. The saliency map is calculated for a given test or reference image (Figures 1(b) and (i)). In our experiments we use the Graph Based Visual Saliency (GBVS) [16], ITTI [24], and the Frequency-Tuned (FT) [25] saliency calculation approaches.
2. The smallest window in the calculated saliency map which has a specific portion of the total salient energy in the image is found (Figures 1(c)-(g) and (j)-(n)). In our experiments we use 20%, 30%, 40%, 50%, 60%, 70%, and 80% of the total salient energy in a given image.
3. The quality of the region detected in step two in the test image will then represent the overall image quality.

Experimental results

To evaluate the performance of the proposed approach we use three benchmark databases, the Tampere Image Database (TID2013) [26], the Computational and Subjective Image Quality (CSIQ) database [4], and the Colourlab Image Database: Image Quality (CID:IQ) [27].

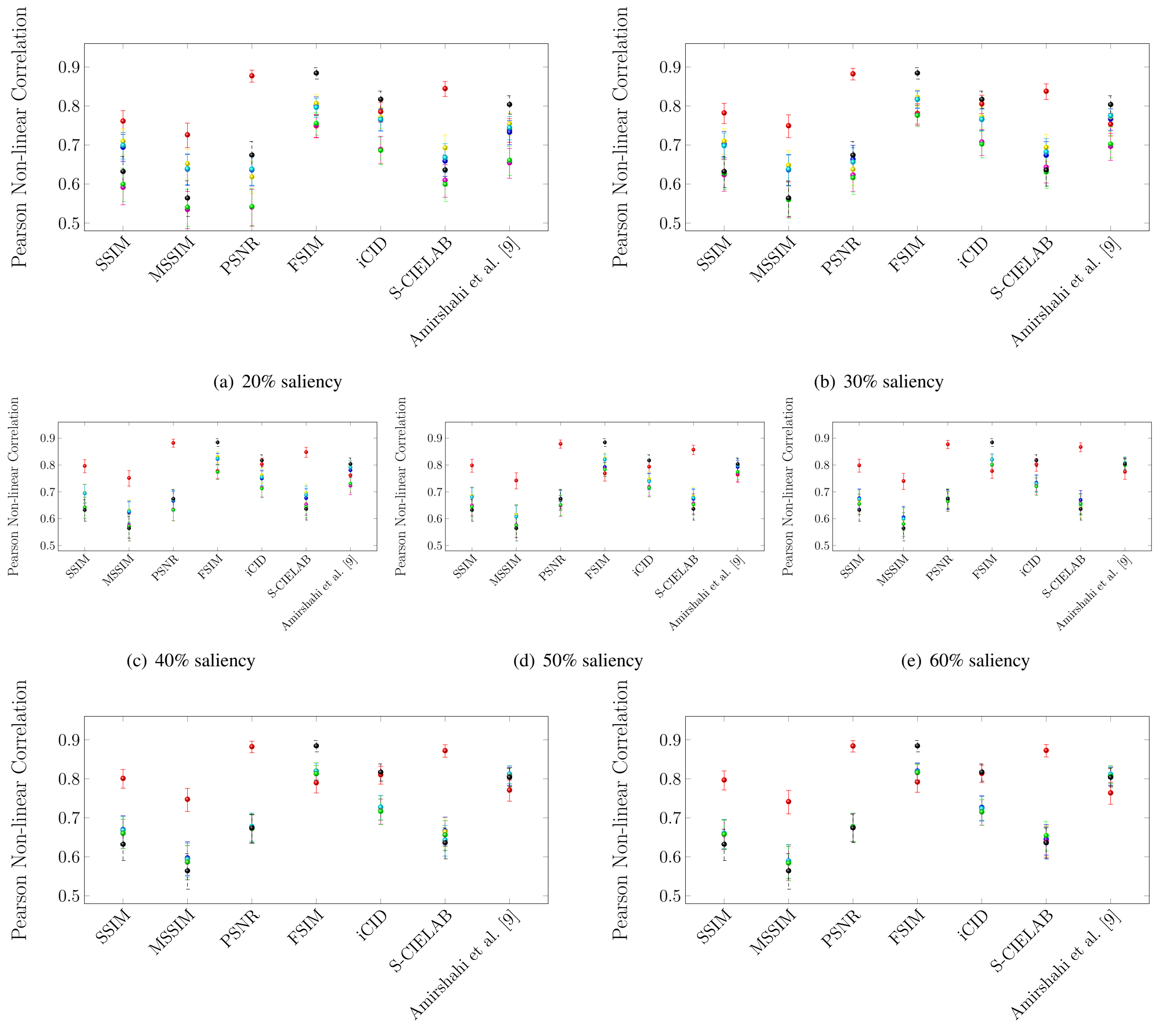
To measure the performance of the proposed approach we calculate the non-linear Pearson correlation between subjective and calculated objective scores. In our experiments we used a wide range of different IQMs. This includes the Structural Similarity Index (SSIM) [5], the Multiscale Structural Similarity Index (MSSIM) [28], the Peak Signal to Noise Ratio (PSNR), the Feature SIMilarity (FSIM) index [29], the iColor-Image-Difference (iCID) [30], S-CIELAB [6], and the CNN based IQM proposed by Amirshahi et al. [9]. Similar to other studies on the performance of IQMs [8], a confidence interval is calculated using Fisher's Z-transform [31], giving us a 95% confidence interval for the correlation values. While we also calculated the Spearman and Kendall correlation coefficients, due to similar correlation rates and space limit we only report the Pearson coefficients. It should be mentioned that the size of the cropped salient region depends on the image and the saliency calculation method used (Table 1).

In the case of the TID2013 database (Figure 2), apart from the FSIM IQM the use of the proposed approach performs better or as good as calculating the quality score for the entire image. Best results are mainly seen in the case of using the GBVS saliency detection method on the reference image while using the FT saliency detection method on test images show the lowest performance. It is interesting to observe that the improvement seen is lower in the case of the IQMs which are based on structural aspects of the image compared to methods such as PSNR which do not pay much attention to this issue. As the amount of salient energy in the cropped region in which we calculate the image quality for increases the difference between the different saliency detection methods and approaches decrease. This issue could basically be linked to the fact that the salient regions detected using different saliency approaches tend to overlap with each other as the amount of salient energy covered increases. It is surprising that the difference between the best performing approach compared to the full image does not change considerably when the amount of salient energy in the cropped region increases from 20% to 80%.

Table 1: Percentage of the image (Figures 1(a) and (h)) covered for a number of different portions of salient energy.

	Salient energy				
	20%	40%	50%	60%	80%
Figure 1(a)	8%	16%	22%	32%	47%
Figure 1(h)	10%	19%	25%	35%	52%

In the case of the CSIQ database (Figure 3), in most cases we observe that the proposed approach is able to improve the accuracy of different IQMs. The ITTI saliency detection methods both calculated for the reference and/or test image shows the best performance among the different possible combinations. With the increase in the amount of salient energy covered in the cropped region, the accuracy of the proposed approach improves to the point that it shows a better accuracy compared to calculating the image quality for the entire image. Using the CNN based IQM proposed by Amirshahi et al. [9], with as low as 30% of the total salient energy, we see better results compared to when the IQM is calculated for the entire image. While as the amount of salient energy in the cropped region increases from 20% to 80% the performance of the proposed approach increases all calculated quality scores show a higher correlation that using the entire image.



(a) 20% saliency (b) 30% saliency

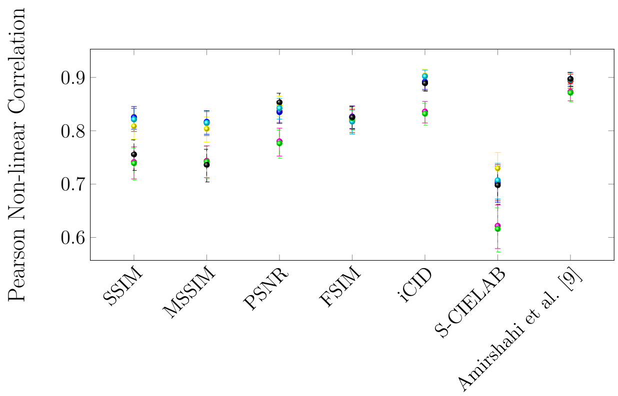
(c) 40% saliency (d) 50% saliency (e) 60% saliency

(f) 70% saliency (g) 80% saliency

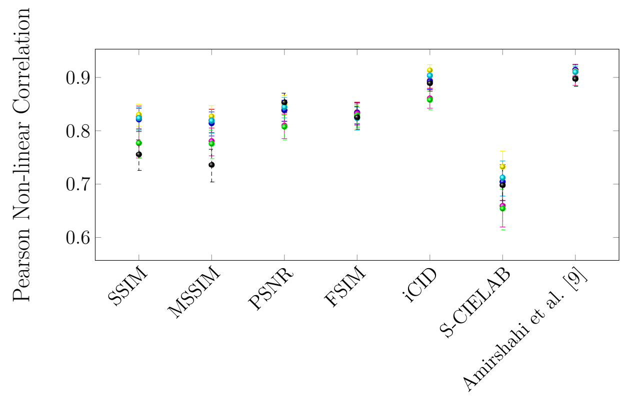
Figure 2. *Non-linear Pearson correlation values for different image quality metrics calculated for the TID2013 dataset for three different saliency calculation methods shown with 95% confidence intervals. In the plots markers in red correspond to the GVBS approach calculated for the reference images, blue correspond to the GVBS approach calculated for the test images, yellow correspond to the ITTI approach calculated for the reference images, cyan correspond to the ITTI approach calculated for the test images, magenta correspond to the FT approach calculated for the reference images, green correspond to the FT approach calculated for the test images, and **black** corresponds to calculating quality of an image in its entire size. Due to space limit we have provided larger figures for two ends of the range. Please zoom in the figure for a better scale of each plot.*

In the case of the CID:IQ database (due to space limits we only show results for the 50cm distance Figure 4), different IQMs in general show a better performance in the case of subjective scores for a viewing distance of 50cm compared to 100cm. This is also the case using the proposed approach where better performance is seen in the case of a viewing distance of 50cm compared to 100cm. While with a viewing distance of 550cm the proposed approach is able to outperform classical image quality evaluation approaches calculated for the entire image, this is not the case for a viewing distance of 100cm. Similar to the CSIQ database the ITTI saliency calculation method provide the best results.

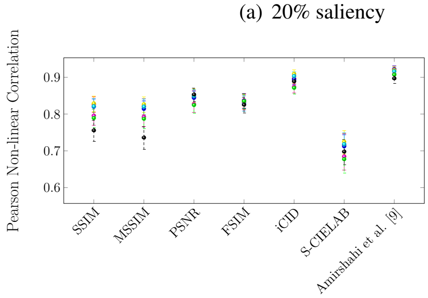
We should point out that the proposed approach showed a low performance in the case of images which had a uniform distribution of saliency across the image (1(h)). In other words, the proposed approach works better when there is a dominant salient region in the image (1(a)). In our experiments we further investigated the performance of the proposed approach on different types of distortions. Our results showed that when compared to the entire image, the proposed approach has the best performance for the Blur, AWGN, and JPEG2000 distortions using the GBVS and ITTI saliency calculation methods. The proposed ap-



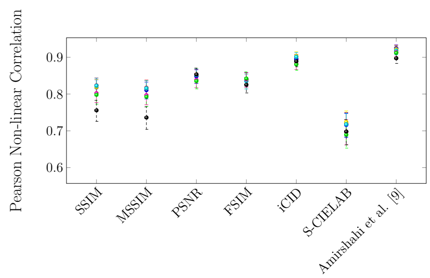
(a) 20% saliency



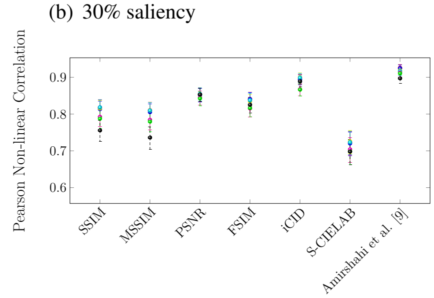
(b) 30% saliency



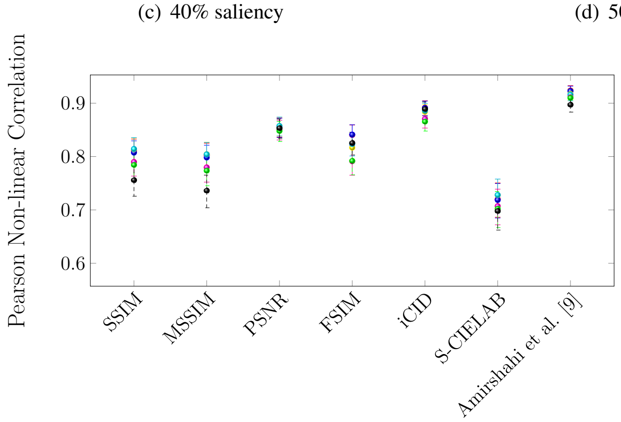
(c) 40% saliency



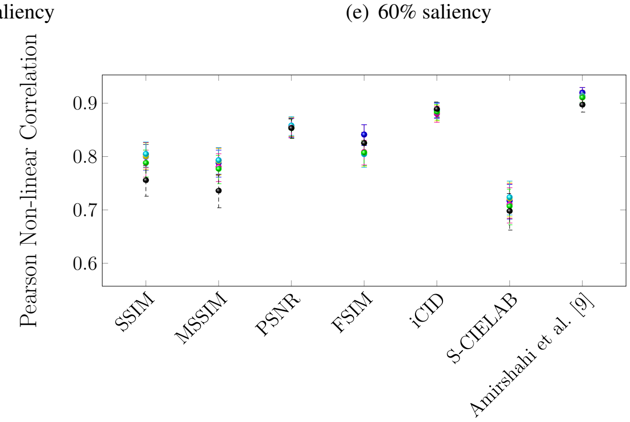
(d) 50% saliency



(e) 60% saliency



(f) 70% saliency



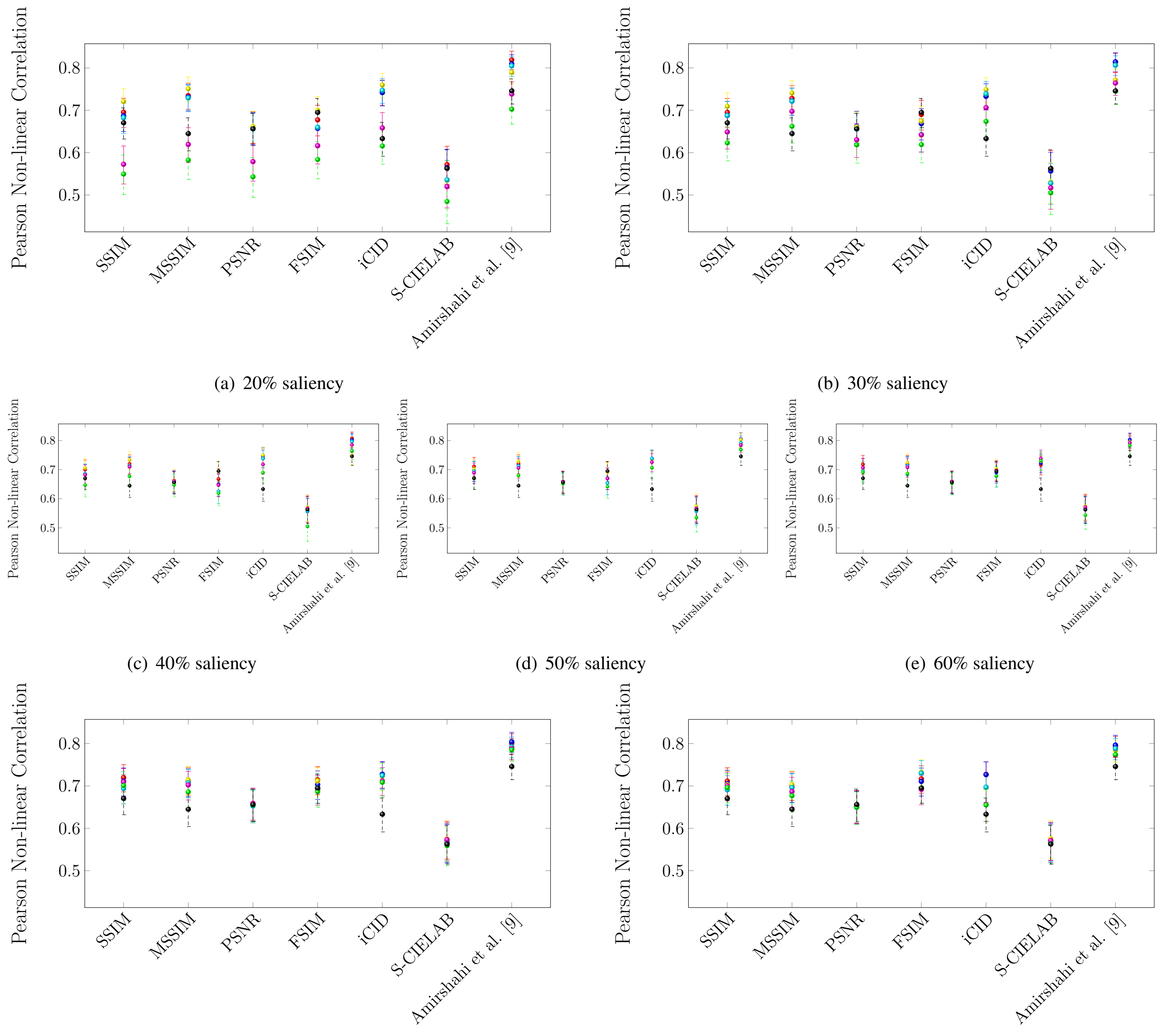
(g) 80% saliency

Figure 3. *Non-linear Pearson correlation values for different image quality metrics calculated for the CSIQ dataset for three different saliency calculation methods shown with 95% confidence intervals. In the plots markers in red correspond to the GVBS approach calculated for the reference images, blue correspond to the GVBS approach calculated for the test images, yellow correspond to the ITTI approach calculated for the reference images, cyan correspond to the ITTI approach calculated for the test images, magenta correspond to the FT approach calculated for the reference images, green correspond to the FT approach calculated for the test images, and* ***black*** *corresponds to calculating quality of an image in its entire size. Due to space limit we have provided larger figures for two ends of the range. Please zoom in the figure for a better scale of each plot.*

proach shows lower collaboration values in the case of FNoise. On average the FT saliency method shows a slight improvement in the case of the Blur, contrast, and JPEG2000 compression distortions while in the case of the JPEG and FNoise distortions using the proposed approach reduces the accuracy of IQMs. In the case of the TID database, the proposed approach shows improvement in JPEG compression, Lossy compression of noisy images, and mean shift. A significant improvement is seen in the CID:IQ dataset with a viewing distance of 50cm in the case of the GBVS and ITTI saliency methods while the FT method does not show a significant change. Other than the JPEG and blur distortions the proposed approach does not show an improvement in the case of the CID:IQ dataset with a viewing distance of 100cm. We should point out that due to the nature of the saliency calculation methods used, the computational time of the proposed approach is shorter than calculating the IQM for the entire image.

Conclusion and future work

In this work, we studied the possibility of just using the quality score calculated for the most salient region in the image to



(a) 20% saliency

(b) 30% saliency

(c) 40% saliency

(d) 50% saliency

(e) 60% saliency

(f) 70% saliency

(g) 80% saliency

Figure 4. *Non-linear Pearson correlation values for different image quality metrics calculated for the CID:IQ dataset with a 50cm viewing distance for three different saliency calculation methods shown with 95% confidence intervals. In the plots markers in red correspond to the GVBS approach calculated for the reference images, blue correspond to the GVBS approach calculated for the test images, yellow correspond to the ITTI approach calculated for the reference images, cyan correspond to the ITTI approach calculated for the test images, magenta correspond to the FT approach calculated for the reference images, green correspond to the FT approach calculated for the test images, and* ***black*** *corresponds to calculating quality of an image in its entire size. Due to space limit we have provided larger figures for two ends of the range. Please zoom in the figure for a better scale of each plot.*

represent the overall quality of the image. For this we find the smallest region in the image that consists of a specific amount of the salient energy of the image. We then calculate the quality of the detected region and use that quality value to represent the quality of the entire image. Results of the experiments on different databases show that the proposed approach has as good if not a better performance compared to the classical approach. In other words, our study shows that "the quality is in the salient region of the image".

References

[1] Seyed Ali Amirshahi. *Aesthetic Quality Assessment of Paintings*. Verlag Dr. Hut, 2015.

[2] Farah Torkamani-Azar and Seyed Ali Amirshahi. A new approach for image quality assessment using svd. In *ISSPA*, pages 1–4, 2007.

[3] Seyed Ali Amirshahi and Marius Pedersen. Future directions in image quality. In *CIC*, pages 399–403, 2019.

[4] Eric C Larson and Damon M Chandler. Most apparent distortion: full-reference image quality assessment and the role

of strategy. *J. Electron. Imaging*, 19(1):011006–011006, 2010.

[5] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, 2004.

[6] Xuemei Zhang, Brian A Wandell, et al. A spatial extension of CIELAB for digital color image reproduction. In *SID international symposium digest of technical papers*, volume 27, pages 731–734. Citeseer, 1996.

[7] Seyed Ali Amirshahi and Farah Torkamani-Azar. Human optic sensitivity computation based on singular value decomposition. *Optica Applicata*, 42(1):137–146, 2012.

[8] Seyed Ali Amirshahi, Marius Pedersen, and Azeddine Beghdadi. Reviving traditional image quality metrics using cnns. In *CIC*, pages 241–246, 2018.

[9] Seyed Ali Amirshahi, Marius Pedersen, and Stella X Yu. Image quality assessment by comparing cnn features between images. *JIST*, 60(6):60410–1, 2016.

[10] Aladine Chetouani, Azeddine Beghdadi, and Mohamed Deriche. Image distortion analysis and classification scheme using a neural approach. In *EUVIP*, pages 183–186, 2010.

[11] Seyed Ali Amirshahi and M-C Larabi. Spatial-temporal video quality metric based on an estimation of qoe. In *QoMEX*, pages 84–89, 2011.

[12] Meisam Jamshidi Seikavandi and Seyed Ali Amirshahi. Evaluating video quality by differentiating between spatial and temporal distortions. In *CVCS*, pages 1–15. CEUR-WS, 2020.

[13] Wei Zhang, Ali Borji, Zhou Wang, Patrick Le Callet, and Hantao Liu. The application of visual saliency models in objective image quality assessment: A statistical evaluation. *IEEE Trans. Neural Netw.*, 27(6):1266–1278, 2015.

[14] Hani Alers, Judith A Redi, Hantao Liu, and Ingrid Heynderickx. Studying the effect of optimizing image quality in salient regions at the expense of background content. *J. Electron. Imaging*, 22(4):043012, 2013.

[15] Seyed Ali Amirshahi and Joachim Denzler. Judging aesthetic quality in paintings based on artistic inspired color features. In *DICTA*, 2017.

[16] Jonathan Harel, Christof Koch, and Pietro Perona. Graph-based visual saliency. In *NIPS*, pages 545–552, 2007.

[17] Jakob Suchan, Mehul Bhatt, Srikrishna Vardarajan, Seyed Ali Amirshahi, and Stella Yu. Semantic analysis of (reflectional) visual symmetry: A human-centred computational model for declarative explainability. *arXiv preprint arXiv:1806.07376*, 2018.

[18] Shao-Fu Xue, Qian Lin, Daniel R Tretter, Seungyon Lee, Zygmunt Pizlo, and Jan Allebach. Investigation of the role of aesthetics in differentiating between photographs taken by amateur and professional photographers. In *Imaging and Printing in a Web 2.0 World III*, volume 8302, page 83020D. International Society for Optics and Photonics, 2012.

[19] Jonas Abeln, Leonie Fresz, Seyed Ali Amirshahi, I Chris McManus, Michael Koch, Helene Kreysa, and Christoph Redies. Preference for well-balanced saliency in details cropped from photographs. *Front. Hum. Neurosci.*, 9:704, 2016.

[20] Long Mai, Hoang Le, Yuzhen Niu, and Feng Liu. Rule of thirds detection from photograph. In *ISM*, pages 91–96, 2011.

[21] Seyed Ali Amirshahi, Gregor Uwe Hayn-Leichsenring, Joachim Denzler, and Christoph Redies. Evaluating the rule of thirds in photographs and paintings. *Art Percept.*, 2(1-2):163–182, 2014.

[22] Judith Redi, Hantao Liu, Rodolfo Zunino, and Ingrid Heynderickx. Interactions of visual attention and quality perception. In *HVEI*, volume 7865, page 78650S, 2011.

[23] Claire Mantel, Nathalie Guyader, Patricia Ladret, Gelu Ionescu, and Thomas Kunlin. Characterizing eye movements during temporal and global quality assessment of h. 264 compressed video sequences. In *HVEI*, volume 8291, page 82910Y, 2012.

[24] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal.*, 20(11):1254–1259, 1998.

[25] Radhakrishna Achanta, Sheila Hemami, Francisco Estrada, and Sabine Susstrunk. Frequency-tuned salient region detection. In *CVPR*, pages 1597–1604, 2009.

[26] Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al. Image database tid2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication*, 30:57–77, 2015.

[27] Xinwei Liu, Marius Pedersen, and Jon Yngve Hardeberg. CID:IQ–a new image quality database. In *ICISP*, pages 193–202. 2014.

[28] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *ACSSC*, volume 2, pages 1398–1402, 2003.

[29] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. Fsim: A feature similarity index for image quality assessment. *IEEE Trans. Image Process.*, 20(8):2378–2386, 2011.

[30] Jens Preiss, Felipe Fernandes, and Philipp Urban. Color-image quality assessment: From prediction to optimization. *IEEE Trans. Image Process.*, 23(3):1366–1378, 2014.

[31] Video Quality Experts Group and others. Final report from the video quality experts group: validation of reduced-reference and no-reference objective models for standard definition television. Technical report, Phase I. Tech. rep.(International Telecommunication Union, Geneva, 2009), 2009.

Author Biography

Meisam Jamshidi Seikavandi is a master graduated from Khaje Nasir University of Technology. He is a Researcher with a demonstrated history of working and research in the Computer Vision area. Recently, he has contributed in several international collaborations.

Seyed Ali Amirshahi is an Associate Professor at the Norwegian University of Science and Technology (NTNU). His research is focused on image quality assessment and computational aesthetics. He received his PhD from the Friedrich Schiller University of Jena in Germany (2015). In 2016 He was a postdoctoral fellow at the International Computer Science Institute (ICSI) in Berkeley, California. From 2017 to 2019 he was employed at NTNU as a FRIPRO/Marie Sklodowska-Curie postdoctoral fellow and a visiting researcher at the University Université Sorbonne Paris Nord.

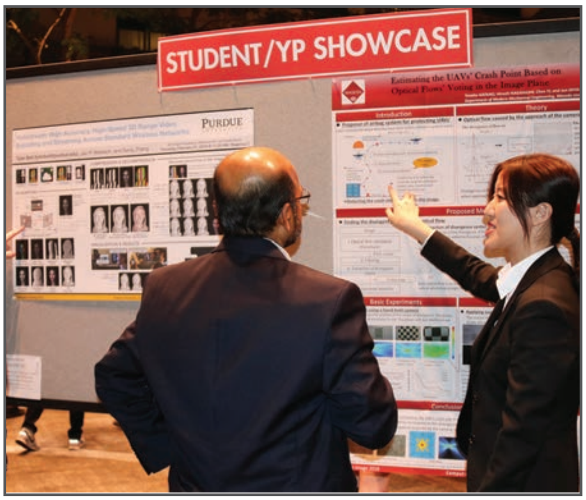
JOIN US AT THE NEXT EI!

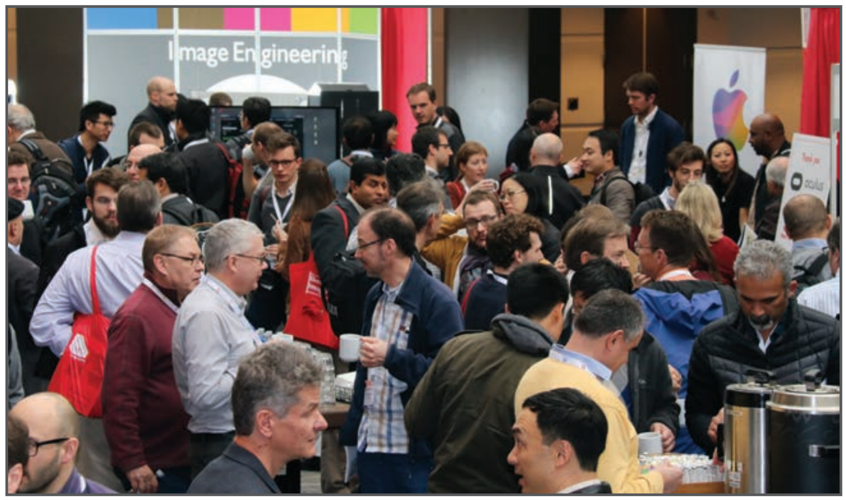
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• **SHORT COURSES** • **EXHIBITS** • **DEMONSTRATION SESSION** • **PLENARY TALKS** •
• **INTERACTIVE PAPER SESSION** • **SPECIAL EVENTS** • **TECHNICAL SESSIONS** •

www.electronicimaging.org

IS&T
imaging.org